Submitted to Anals of Mathematical Statistics, Nov. 1968

A Geometric Approach to Multivariate Analysis - 1. Linear Estimation

Duke University Grant NGL-34-001-005

Summary

This research is concerned with the general multivariate linear model from a vector space point of view. That is the usual n x p "data matrix" is treated as a vector thereby permitting the analysis to be visualized in the same way that it is in the univariate case. The difference being that the fundamental building blocks of the estimation space are p-dimensional subspaces instead of the univariate cases's 1-dimensional ones. The present paper establishes this point of view and applies it to the problem of linear estimation. In doing so we find for example that S. N. Roy's statement of the multivariate linear hypothesis is not the most general "estimable" hypothesis. Work on appropriate tests of hypothesis is underway.

A Geometric Approach to Multivariate Analysis - 1. Linear Estimation

0. Introduction. This research is concerned with the general multivariate linear model which is treated extensively in Anderson [1] and Roy [3]. The principal focus of attention in this work is on the vector space or geometrical interpretation of linear unbiased estimates of linear combinations of the parameters in the model. Such interpretation of the univariate linear model has proved enlightening in the analysis and theoretical development of univariate analysis of variance, e.g. Bose [2], Scheffe [4]. It is unlikely that any of the results in this paper are new per se, but rather that the setting is new.

In univariate analysis of variance the geometrical view seems especially important when making a postiori inferences from the data which are not precise enough to justify the specification of a P-value. Such "detective work", as Tukey calls it, is greatly enhanced by being able to visualize the variation due to a certain linear combination of parameters as the squared length of the projection of the observation vector on a certain subspace. An analogous geometric view for the multivariate case would be useful and it is the object of this work to provide the requisite preliminaries for such a view. More work along these lines is in progress.

Minimum variance, unbiased, linear estimates (often called best estimates) of linear combinations of design parameters in the univariate case are defined by certain one dimensional subspaces of the estimation space which is in turn defined by the design matrix. This result is useful in "visualizing" the estimates and in defining the sums of squares due to linear combinations of the parameters. In this paper it is shown that in the p-variate case "best" estimates of linear combinations of certain parameter vectors are defined by p-dimensional subspaces of the estimation space which is itself defined by a suitably modified design matrix. Then linear combinations of the parameters are considered as linear combinations of the components of linear combinations of the parameter vectors and their best estimates obtained. Predictably, these best estimates turn out to be defined by certain one dimensional subspaces of the estimation space. It then follows that the fundamental building blocks of the estimation space in the p-variate case are certain p-dimensional subspaces - a natural extension of the univariate case. Furthermore orthogonal generating sets for these subspaces are readily available from the design matrix.

In the final sections the straight forward generalizations of the usual results on the normal equations are given together with an appropriate statement of the Gauss-Markoff theorem.

A comment here on notation will prove helpful. In what follows the vector space generated by the rows (columns) of a matrix A will be denoted by $V_r(A)$ ($V_c(A)$). Thus, for example, $V_c(A) = V_r(A')$.

1. <u>Multivariate General Linear Model.</u> Let $\underline{x}_1, \ldots, \underline{x}_N$ be N independent, identically distributed, p-dimensional, random vectors with real coordinates. By the assumption that these random vectors obey the general linear model we shall mean that there exists a known, constant $N \times k$ (design) matrix A_0, an unknown, constant $k \times p$ (parameter) matrix Θ, and a random $N \times p$ (error) matrix F so that

$$(1.1) \qquad \begin{pmatrix} \underline{x}'_1 \\ \underline{x}'_2 \\ \vdots \\ \underline{x}'_N \end{pmatrix} = X = A_0 \Theta + F, \quad E(F) = 0 .$$

It is with such a model that both Anderson and Roy are concerned in their respective approaches to estimating Θ and multivariate analysis of variance. We shall consider a different approach to this model and for this approach a different, but equivalent, formulation of (1.1) is needed. Let $A_0 = (a_{ij})$; let $\underline{\mu}'_1, \ldots, \underline{\mu}'_k$ be the $1 \times p$ vectors which form the rows of Θ; and let $\underline{e}'_1, \ldots, \underline{e}'_N$ be the $1 \times p$ vectors which form the rows of F. If we let I_p be the $p \times p$ identity matrix, then the assumptions embodied in (1.1) can be written equivalently as

$$(1.2) \qquad x = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_N \end{bmatrix} = \begin{bmatrix} a_{11} I_p & a_{12} I_p \cdots & a_{1k} I_p \\ a_{21} I_p & a_{22} I_p \cdots & a_{2k} I_p \\ \vdots & \vdots & \vdots \\ a_{N1} I_p & a_{N2} I_p \cdots & a_{Nk} I_p \end{bmatrix} \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \\ \vdots \\ \underline{\mu}_k \end{bmatrix} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_N \end{bmatrix}$$

where the usual notation for partitioned matrices has been employed. Making the obvious identifications we get that (1.2) can be written as

$$(1.3) \qquad \underset{Np \times 1}{\underline{x}} = \underset{Np \times kp}{A} \; \underset{kp \times 1}{\underline{\mu}} + \underset{Np \times 1}{\underline{e}}, \quad E(\underline{e}) = 0.$$

Before proceeding we shall define two operators which will enable us to move more gracefully between the formulation of the model in (1.1) and that in (1.2).

Definition 1.1 (a) Let $A_{m \times n}$ be an $m \times n$ real matrix with row vectors $\underline{a}'_1, \ldots, \underline{a}'_m$. Define the "stretch" operator $\mathcal{S}$ by

$$(1.4) \qquad \mathcal{S}(A_{m \times n}) = \begin{bmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_m \end{bmatrix}$$

(b) Let $B_{r \times s}$ be an $r \times s$ real matrix and define the "insertion" operator $\mathcal{O}$ (of B into A) by

$$(1.5) \qquad \mathcal{O}(B_{r \times s} : A_{m \times n}) = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & & & \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

where $A_{m \times n} = (a_{ij})$.

This operator is a form of the direct product defined for example in Wedderburn [5]. Routine calcutions are sufficient to establish the following basic properties of the "stretch and "insertion" operators. Here $A_{m \times n}$, $B_{n \times q}$, $C_{r \times s}$, $D_{s \times u}$, $E_{s \times u}$ are real matrices of the size indicated and I_r is the $r \times r$ identity.

(i) $\mathcal{S}(D_{s \times u} + E_{s \times u}) = \mathcal{S}(D_{s \times u}) + \mathcal{S}(E_{s \times u})$

(ii) $\mathcal{S}(A_{m \times n} B_{n \times p}) = \mathcal{O}(I_p : A_{m \times n}) \mathcal{S}(B_{n \times p})$

(iii) $\mathcal{O}(C_{r \times s} : A_{m \times n}) \ \mathcal{O}(D_{s \times u} : B_{n \times q}) = \mathcal{O}(CD : AB)$

(iv) $A_{m \times n} = \mathcal{O}(A_{m \times n} : I_1)$, $\mathcal{O}(B : A)' = \mathcal{O}(B' : A')$

where for any matrix M, M' is its transpose.

We can now write (1.3) as

(1.6) $\mathcal{S}(X) = \mathcal{O}(I_p : A_0) \ \mathcal{S}(\Theta) + \mathcal{S}(F)$, $E[\mathcal{S}(F)] = 0$

so that we have an $Np \times 1$ observation vector $\underline{x} = \mathcal{S}(X)$, an $Np \times kp$ design matrix $A = \mathcal{O}(I_p : A_0)$, a $kp \times 1$ parameter vector $\underline{\mu} = \mathcal{S}(\Theta)$, and an $Np \times 1$ error vector $\underline{e} = \mathcal{S}(F)$. Using these definitions and the properties of the "insertion" and "stretch" operators we can easily make the transition from (1.1) to (1.3) and back.

If C' is an $s \times Np$ matrix, then

(1.7) $E(C'\underline{x}) = C' A \underline{\mu}$.

If we further assume that $\underline{x}_1$ has a finite $p \times p$ variance-covariance matrix Σ_0, then the variance-covariance matrix of $\underline{x}$ (denoted by VAR $\underline{x}$) is given by an $Np \times Np$ real, symmetric,positive definite matrix Σ defined by

$$(1.8) \qquad \Sigma = \text{VAR } \underline{x} = \mathcal{O}(\Sigma_0 : I_N)$$

and

$$(1.9) \qquad \text{VAR } C' \underline{x} = C' \Sigma C.$$

The general linear model for the vectors $\underline{x}_1, \ldots, \underline{x}_N$ has now been put in a form which encourages the use of vector space ideas in the analysis of this model.

2. Estimable linear functions of $\underline{\mu}_1, \ldots, \underline{\mu}_k$. In multivariate analysis of variance one often wishes to test hypotheses of the form $L_0 \Theta = 0$ where L_0 is a $g \times k$ matrix. Thus estimation of $L_0 \Theta$ might naturally arise. Note that $L_0 \Theta$ is simply g linear combinations of $\underline{\mu}_1, \ldots, \underline{\mu}_k$. Thus we shall consider the problem of finding an unbiased estimate of a linear combination

$$\ell_1 \underline{\mu}_1 + \ell_2 \underline{\mu}_2 + \cdots + \ell_k \underline{\mu}_k$$

for arbitary real numbers $\ell_1, \ell_2, \ldots, \ell_k$. We must find out which, if any, such linear combinations have unbiased estimates ; and if an unbiased estimate exists, what it is.

Let the $p \times kp$ matrix L' be defined by

$$(2.1) \qquad L' = (\ell_1 I_p, \ \ell_2 I_p, \ldots, \ell_k I_k) = \mathcal{O}(I_p : \underset{\sim}{\ell}')$$

for $\underset{\sim}{\ell}' = (\ell_1, \ldots, \ell_k)$ and note that

$$(2.2) \qquad L' \underline{\mu} = \ell_1 \underline{\mu}_1 + \ell_2 \underline{\mu}_2 + \ldots + \ell_k \underline{\mu}_k .$$

In our search for unbiased estimates of $L'\underline{\mu}$ we shall restrict our attention to linear functions of $\underline{x}$. Thus we only consider functions of the form $C'\underline{x}$ where C is an $Np \times p$ real matrix.

Definition 2.1 An $m \times 1$ vector $\underline{\theta}$ will be said to be estimable with respect to $\underline{x}$ if there exists a linear function of $\underline{x}$ which is an unbiased estimate of $\underline{\theta}$.

Actually when considering $L'\underline{\mu}$ we can be more specific about the linear function of $\underline{x}$ in question.

Lemma 2.1 The linear function $L'\underline{\mu}$ is estimable if and only if there exists N real numbers $c_1, \ldots, c_N$ so that $E(C'\underline{x}) = L'\underline{\mu}$ for C defined by

$$C' = [c_1 I_p, c_2 I_p, \ldots c_N I_p] = \otimes(I_p : \underline{c}'). \tag{2.3}$$

Proof: A general linear function of $\underline{x}$ appropriate for estimating $L'\underline{\mu}$ would be $C'\underline{x}$ for C an $Np \times p$ matrix. Requiring such an estimate to be unbiased means that $C'A = L'$. Routine considerations of this equation show that one need only consider C of the form $\otimes(I_p : \underline{c})$ for $\underline{c}$ an $N \times 1$ real vector.

Since $C'\underline{x}$ for C as in (2.3) is simply a linear combination of $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N$, viz. $c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_N\underline{x}_N$, the lemma is both expected and welcome. From lemma 2.1 we see that the following theorem is true.

Theorem 2.1 *A necessary and sufficient condition for* $L'\underline{\mu}$ *to be estimable is for* $(\ell_1, \ldots, \ell_k)$ *to be a linear combination of the rows of* A_0. *Or in terms of (1.2)* $L'\underline{\mu}$ *is estimable if and only if* L' *is a linear combination of* $\mathcal{O}(I_p : \underline{a}'_1)$, $\mathcal{O}(I_p : \underline{a}'_2), \ldots, \mathcal{O}(I_p : \underline{a}'_N)$ *for* $\underline{a}'_1, \ldots, \underline{a}'_N$ *the rows of* A_0.

Proof: From the lemma we have that estimability is equivalent to the existance of a real vector $\underline{c}$ so that $\mathcal{O}(I_p : \underline{c}')\ \mathcal{O}(I_p : A_0) = \mathcal{O}(I_p : \underline{\ell}')$ which implies $\underline{c}'A_0 = \underline{\ell}'$.

Notice that the rows of a matrix of the form

$$(2.4) \qquad [a_1 I_p, a_2 I_p, \ldots, a_q I_p] = \mathcal{O}(I_p : \underline{a}') \text{ for } \underline{a}' = (a_1, \ldots, a_q)$$

form an orthogonal set of qp-dimensional real vectors which generate a p-dimensional subspace of the qp-dimensional real vector space R^{qp}. Similarly for the column vectors of the transpose of the matrix in (2.4). Thus we make the following definition.

Definition 2.2 *If* $\underline{a}$ *is a* $q \times 1$ *real vector, then we shall call* $\mathcal{O}(I_p : \underline{a})$ *a p-dimensional orthogonal generating set in* R^{qp}. *The vector* $\underline{a}$ *will be called the defining vector of* $\mathcal{O}(I_p : \underline{a})$. *When the dimensions are understood from context we shall simply use the abbreviation o.g.s. for an orthogonal generating set. We shall speak of submatrices of the form of (2.4) which form rows (columns) of a matrix as row (column) p-dimensional o.g.s. in* R^{qp}.

We shall view a p-dimensional o.g.s. in R^{qp} much as we do a 1-dimensional vector in R^q. It is clear that for $p = 1$ an o.g.s. is a vector which in turn is a generating set for a 1-dimensional subspace of R^q. Definitions of orthogonality, independence and equivalence of o.g.s. are similar to those for vectors and are formally stated below. In order to avoid awkward notation the plural of o.g.s. will also be written o.g.s. and the number inferred from context.

Definition 2.3 (a) The vector space generated by a set of r o.g.s. is the space generated by the rp qp-dimensional vectors which make up the rows of the r o.g.s.

(b) Two o.g.s. will be said to be equivalent if they generate the same subspace of R^{qp}.

(c) Two o.g.s. will be said to be orthogonal if the corresponding q-dimensional defining vectors are orthogonal. A set of o.g.s. will be called an orthogonal set if every two distinct o.g.s. are orthogonal.

(d) A set of o.g.s. will be said to be linearly independent if the corresponding set of defining vectors is linearly independent.

Lemma 2.2 If V is a subspace of R^{qp} generated by k p-dimensional o.g.s. $\mathcal{O}(I_p : \underline{a}_1), \ldots, \mathcal{O}(I_p : \underline{a}_k)$ then the space generated by a p-dimensional o.g.s. $\mathcal{O}(I_p : \underline{c})$ is contained in V if and only if there exist real numbers $d_1, \ldots, d_k$ such that

(2.6) $\underline{c} = d_1 \underline{a}_1 + \ldots + d_k \underline{a}_k$.

Proof: The sufficiency is clear. The necessity follows by considering the orthogonal complement to the space generated by $\{\underline{a}_1, \ldots, \underline{a}_k\}$; and the fact that if $\underline{a}$, $\underline{b}$ are orthogonal vectors in R^q, then $\underline{m}'_1 \otimes (I_p : \underline{a}')$ and $\underline{m}'_2 \otimes (I_p : \underline{b}')$ are orthogonal vectors in R^{qp} for $\underline{m}_1$ and $\underline{m}_2$ arbitrary $p \times 1$ vectors.

Note: This means that if the subspace generated by an o.g.s. C is in V, $C = d_1 A_1 + d_2 A_2 + \ldots + d_k A_k$ for $A_i = \otimes(I_p : \underline{a}_i)$ $i = 1, \ldots, k$.

In this context we can restate the result of Theorem 2.1 as $L'\underline{\mu}$ is estimable if and only if L' generates a subspace of R^{kp} contained in the subspace generated by the row p-dimensional o.g.s. in R^{kp} of A, i.e. $V_c(L) \subset V_r(A)$.

Definition 2.4 The subspace of R^{kp} defined by the row p-dimensional o.g.s. of A will be called the parameter space for the model (1.2) and denoted by $V_r(A)$. Notice that the notational convention established in the introduction and definition 2.3 (a) dictate this notation.

Lemma 2.3 If the rank of A_0 is $s \leq k < N$, the rank of A is sp and the dimension of $V_r(A)$ is sp.

Proof: The result follows immediately from the definitions of A and $V_r(A)$.

We have thus determined which linear functions $\underline{\mu}_1, \ldots, \underline{\mu}_k$ have linear unbiased estimates. We shall have more to say about these estimates in section 4.

3. Estimable functions of the components of $\underline{\mu}$. In section 2 we concentrated on linear functions of $\underline{\mu}_1, \underline{\mu}_2, \ldots, \underline{\mu}_k$. This led to investigation of certain p-dimensional subspaces of the parameter space, $V_r(A)$. In this section we shall be interested in certain 1-dimensional subspaces of $V_r(A)$ as we shall be interested in estimating linear functions of the (scalar) components of $\underline{\mu}$ as opposed to the (vector) components $\underline{\mu}_1, \ldots, \underline{\mu}_k$.

Our study in section 2 was motivated by the consideration of the hypothesis $L_0 \Theta = 0$. A more general hypothesis which suggests estimation problems of the form mentioned in this section is

$$(3.1) \qquad \underset{g \times k}{B} \quad \underset{k \times p}{\Theta} \quad \underset{p \times u}{M} = \underset{g \times u}{0}$$

where rank of B (rk B) is $g \leq \text{rk } A_0$ and $\text{rk } M = u \leq p$.

If $\underline{b}'_i = (b_{i1}, \ldots, b_{ik})$ is ith row of B and $\underline{m}_j$ is jth column of M, then (3.1) can be stated as follows

$$(3.2) \qquad \underset{gu \times kp}{Q} \quad \underset{kp \times 1}{\underline{\mu}} = \underset{gu \times 1}{\underline{0}}$$

where Q is a matrix whose gu rows are given (in any order)by

$$(3.3) \qquad \underline{q}'_{ij} = (b_{i1}\ \underline{m}'_j,\ b_{i2}\ \underline{m}'_j, \ldots, b_{ik}\ \underline{m}'_j) = \otimes(\underline{m}'_j : \underline{b}'_i)$$

for $i = 1 \ldots g$, $j = 1, \ldots, u$. We have the following lemma.

Lemma 3.1 If Q is defined as in (3.3) and B and M are defined as in (3.1),

$$(3.4) \qquad \text{rk}\, Q = (\text{rk } B)(\text{rk } M).$$

Proof: The result follows from a straight forward analysis of the equation $\sum_{i=1}^{g} \sum_{j=1}^{u} \gamma_{ij} \underline{q}_{ij} = \underline{0}$ using the fact that $\underline{q}_{ij} = \otimes (\underline{m}_j : \underline{b}_i)$.

It should be noted that if Q is an arbitrary gu x kp matrix of rank gu, then there do not necessarily exist matrices B and M so that (3.2) implies (3.1).

Let us consider an arbitrary row of Q, $\underline{q}'_{ij}$, and investigate the problem of finding a linear unbiased estimate of $\underline{q}'_{ij} \underline{\mu}$ - a linear function of the components of $\underline{\mu}$. For convenience we consider $\underline{q}'_{11} \underline{\mu}$. Now

$$\underline{q}'_{11} \underline{\mu} = \otimes (\underline{m}'_1 : \underline{b}'_1) \underline{\mu} = \otimes (\underline{m}'_1 : I_1) \otimes (I_p : \underline{b}'_1) \underline{\mu} = \underline{m}'_1 L'_1 \underline{\mu}$$

where L'_1 is the o.g.s. given by $\otimes (I_p : \underline{b}'_1)$. Thus $\underline{q}'_{11} \underline{\mu}$ is a linear function of the components of a linear function of $\underline{\mu}_1, \ldots, \underline{\mu}_k$.

If $L'_1 \underline{\mu}$ is estimable, there exists a p-dimensional o.g.s. C_1 in $R^{N}p$ so that $E(C'_1 \underline{x}) = L'_1 \underline{\mu}$. Thus $E(\underline{m}'_1 C'_1 \underline{x}) = \underline{m}'_1 L'_1 \underline{\mu}$, i.e. $m'_1 C'_1 \underline{x}$ is a linear unbiased estimate of $\underline{q}'_{11} \underline{\mu}$. Note that $\underline{m}'_1 C'_1$ is a 1 x Np vector so that $\underline{m}'_1 C'_1 \underline{x}$ is of the form $\underline{c}' \underline{x}$ and thus $\underline{q}'_{11} \underline{\mu}$ is estimable.

Theorem 3.1 Let $\underline{m}$ be a p x 1 non-zero real vector and let L be a p-dimensional o.g.s. in R^{kp}. Then a necessary and sufficient condition for $\underline{m}' \underline{L}' \underline{\mu}$ to be estimable is that $L' \underline{\mu}$ be estimable. Furthermore if $L' \underline{\mu}$ is estimable, then there exists a p-dimensional o.g.s. in R^{Np}, say C, such that $\underline{m}' C' \underline{x}$ is a linear unbiased estimate of $\underline{m}' L' \underline{\mu}$.

Proof: The sufficiency is clear. For the necessity, if $\underline{m}'$ L' $\underline{\mu}$ is estimable, then there is an $Np \times 1$ real vector $\underline{c}$ so that $\underline{c}'\,\underline{x}$ is an unbiased estimate of $\underline{m}'$ L' $\underline{\mu}$. This results in $\underline{c}'\,A = \underline{m}'\,L'$; and if $\underline{b}$ is such that $L = \bigcirc(I_p : \underline{b})$, it follows that for $\underline{m} \neq \underline{0}$, $\underline{b}$ is in the space $V_r(A_0)$. Thus $V_c(L) \subset V_r(A)$ and the proof is complete. Thus $\underline{q}'_{ij}\,\underline{\mu}$ is estimable if and only if the ith row of B belongs to $V_r(A_0)$.

Notice that $\underline{m}'$ L' generates a 1-dimensional subspace of $V_c(L)$; in fact, every 1-dimensional subspace of $V_c(L)$ is generated by $\underline{m}'$ L' for some vector $\underline{m}$. Furthermore if rk $A = kp$, then for every o.g.s. L, $L'\,\underline{\mu}$ is estimable ; and then every $\underline{m}'$ L' $\underline{\mu}$ has a linear unbiased estimate. Recall that rk $A = kp$ implies dimension of $V_r(A)$ (dim $V_r(A)$) is kp. More generally suppose dim $V_r(A) = sp$, $s \leq k$ and suppose $L_1, L_2, \ldots, L_s$ are s independent p-dimensional o.g.s. in R^{kp} each generating a subspace of $V_r(A)$. Then any vector in $V_r(A)$ can be written as

$$(3.6) \qquad \underline{m}'_1\, L'_1 + \underline{m}'_2\, L'_2 + \ldots + \underline{m}'_s\, L'_s$$

for a unique choice of the s $p \times 1$ vectors $\underline{m}_1, \ldots, \underline{m}_s$.

Thus we have the following theorem.

Theorem 3.2 *If $\underline{\ell} \in V_r(A)$, there exists a vector $\underline{c} \in R^{Np}$ such that $\underline{c}'\,\underline{x}$ is a linear unbiased estimate of $\underline{\ell}'\,\underline{\mu}$. Furthermore since $\underline{\ell}$ can be written as in (3.6), a choice for $\underline{c}$ would be*

(3.7) $\underline{c}' = \underline{m}'_1 C'_1 + \underline{m}'_2 C'_2 + \ldots + \underline{m}'_s C'_s$

where C_i is a p-dimensional o.g.s. in R^{Np} such that $E(C'_i \underline{x}) = L'_i \underline{\mu}$ for $i = 1, \ldots, s$.

Let us now consider the hypothesis (3.1) or equivalently (3.2) in light of theorems 3.1 and 3.2. The hypothesis (3.2) requires that in the estimable case certain one dimension subspaces of $V_r(A)$ be orthogonal to $\underline{\mu}$. Furthermore these one dimensional spaces are constrained to lie in $V_c(L)$ for certain o.g.s. L. Notice that all one dimensional subspaces in $V_r(A)$ are not of this form, e.g. if $k = p = 2$, $(1, 0, 0, 1)$ does not lie in a subspace of $V_c(L)$ for any o.g.s. L. Theorem 3.2 shows that more general linear hypotheses can be considered without sacrificing estimability. In fact, there is a one to one correspondence between linear, estimable hypotheses and the one dimensional subspaces of $V_r(A)$.

Notice the relationship between the vector $\underline{\ell}$ in the parameter space $V_r(A)$ as given in (3.6) and its "estimate" as given in (3.7). In section 4 we shall make this relationship even more specific.

4. The fundamental theorem of linear estimation. In this section we shall consider the problem of finding the "best" estimate of an estimable linear function of $\underline{\mu}_1, \ldots, \underline{\mu}_k$, i.e. finding the "best" estimate of $L' \underline{\mu}$ for L, a p-dimensional o.g.s. in R^{kp} such that $V_c(L) \subset V_r(A)$. We now state a lemma which will be useful in developing the idea of "best" estimate.

Lemma 4.1 Let C be a p-dimensional o.g.s. in R^{Np} and let $\underline{c}$ be the $N \times 1$ defining vector for C. Then

$$(4.1) \qquad \text{VAR}\,(C'\,\underline{x}) = C'\,C\,\Sigma_0 = (\underline{c}'\,\underline{c})\,\Sigma_0 \,.$$

We shall call the non-negative number $\underline{c}'\,\underline{c}$ associated with the o.g.s. C the covariance coefficient of $C'\,\underline{x}$.

Proof: $\text{VAR}\,(C'\,\underline{x}) = C'\,\Sigma\,C$
$$\begin{aligned} &= \mathcal{O}(I_p : \underline{c}')\,\mathcal{O}(\Sigma_0 : I_p)\,\mathcal{O}(I_p : \underline{c}) \\ &= \mathcal{O}(I_p : \underline{c}')\,\mathcal{O}(I_p : \underline{c})\,\mathcal{O}(\Sigma_0, I_1) \\ &= C'\,C\,\Sigma_0 = \underline{c}',\,\underline{c}\,\Sigma_0 \end{aligned}$$

Thus we make the following definition.

Definition 4.1 A best o.g.s. estimate of $L'\,\underline{\mu}$ will be an estimate of the form $C'\,\underline{x}$ for C a p-dimensional o.g.s. in R^{Np} which satisfies:

(i) $E(C'\,\underline{x}) = L'\,\underline{\mu}$

(ii) The covariance coefficient of $C'\,\underline{x}$ is at most that of any other o.g.s. estimate satisfying (i).

In other words a best o.g.s. estimate of a linear combination of $\underline{\mu}_1, \ldots, \underline{\mu}_k$ is a linear function of $\underline{x}_1, \ldots \underline{x}_N$ which is unbiased and has minimum covariance coefficient.

We shall now set about showing that there is a unique best o.g.s. estimate for an estimable $L'\,\underline{\mu}$. Toward this end we shall decompose the observation space R^{Np} into two orthogonal subspaces.

Definition 4.2 Let $V_c(A)$ be the subspace of R^{Np} generated by the k column p-dimensional o.g.s. of A. We shall call $V_c(A)$ the estimation space. Let $V(E)$ be the subspace of R^{Np} generated by $\{C: E(C' \underline{x}) = 0,$ C a p-dimensional o.g.s. in $R^{Np}\}$. We shall call $V(E)$ the error space.

We shall abuse the notation by saying an o.g.s. belongs to $V_c(A)$, for example, when we mean the subspace generated by the o.g.s. is contained in $V_c(A)$.

By using the fact that the defining vectors of the o.g.s. which generate $V_c(A)$ and $V(E)$ generate s and $N - s$ dimensional subspaces of R^N respectively and that R^N is the direct sum of these subspaces we can establish the next lemma. This lemma may also be proven by considering the N-dimensional vector space of o.g.s. over the field consisting of elements of the form aI_p, $a \in R$. Here $s = \text{rk } A_0$.

Lemma 4.2 (i) The estimation space $V_c(A)$ is orthogonal to the error space $V(E)$ and R^{Np} is the direct sum of $V_c(A)$ and $V(E)$.

(ii) If rk $A = sp \leq kp$, then dimension $V_c(A) = sp$ and dimension $V(E) = (N - s)p$.

(iii) $V(E) = \{\underline{c} \in R^{Np}: E(\underline{c}' \underline{x}) = 0\}$

(iv) Given a p-dimensional o.g.s. D in R^{Np} there exist unique p-dimensional o.g.s. C and F so that $D = C + F$, C belongs to $V_c(A)$, F belongs to $V(E)$, and $V_c(C)$ is orthogonal to $V_c(F)$.

We are now in a position to state a theorem which locates the unique best o.g.s. estimate of an estimable linear combination of $\underline{\mu}_1, \ldots, \underline{\mu}_k$ in the space $V_c(A)$.

Theorem 4.1 (Fundamental Theorem of Linear Estimation).

If $L'\underline{\mu}$ is an estimable linear function of $\underline{\mu}_1, \ldots, \underline{\mu}_k$ then there exists a unique o.g.s. C in $V_c(A)$ such that $C'\underline{x}$ is the best o.g.s. estimate of $L'\underline{\mu}$.

Proof: Since $L'\underline{\mu}$ is estimable, there exists an o.g.s. D in R^{Np} so that $E(D'\underline{x}) = L'\underline{\mu}$. By lemma 4.2 (iv), there exist unique o.g.s. C in $V_c(A)$ and F in $V(E)$ such that $D = C + F$. Then since $EC'\underline{x} = L'\underline{\mu}$ and $VAR(D'\underline{x}) = (C' + F')(C + F)\Sigma_0 = C'C\Sigma_0 + F'F\Sigma_0 \geq VAR(C'\underline{x})$, it is clear that $C'\underline{x}$ is the unique best o.g.s. estimate of $L'\underline{\mu}$.

Corollary 4.1.1 If $L'\underline{\mu}$ is an estimable linear function of $\underline{\mu}_1, \ldots, \underline{\mu}_k$; there exists a p-dimensional o.g.s. in R^{kp}, Q, so that $Q'A'\underline{x}$ is the best o.g.s. estimate of $L'\underline{\mu}$.

Proof: Since C is in $V_c(A)$, lemma 2.2 implies that there exist real numbers $q_1, \ldots, q_k$ so that $C = q_1 A_1 + \ldots + q_k A_k$ where A_i are the column o.g.s. of A. The result follows by defining $Q = \circledS(I_p : \underline{q})$ for $\underline{q}' = (q_1, \ldots, q_k)$.

Since the idea of a best estimate of a real parameter is usually associated with minimum variance, linear, unbiased estimates, we state the following definition generalizing this idea to parameter vectors.

Definition 4.3 A best estimate of an $m \times 1$ vector $\underline{\theta}$ based on an $M \times 1$ observation vector $\underline{x}$ will be any linear function of $\underline{x}$ which is unbiased and whose components have minimum variance.

Now we have the following corollary to Theorem 4.1 .

Corollary 4.1.2 The best o.g.s. estimate of an estimable linear function $L' \underline{\mu}$ is the best estimate of $L' \underline{\mu}$.

Notice that in general the condition of having minimum covariance coefficient is stronger than that of having minimum component variance in that the former places constraints on the covariances while the latter does not.

The case of linear functions of the components of $\underline{\mu}$ can now be viewed in the light of the decomposition of R^{Np}, and the following improvement of theorem 3.2 can be obtained.

Theorem 4.2 If $\underline{\ell} \in V_r(A)$, there exists a unique vector $\underline{c} \in V_c(A)$ such that $\underline{c}' \underline{x}$ is the best estimate of $\underline{\ell}' \underline{\mu}$, i.e.

(i) $E(\underline{c}' \underline{x}) = \underline{\ell}' \underline{\mu}$

(ii) $\mathrm{Var}(\underline{c}' \underline{x})$ is less than the variance of any other linear function of $\underline{x}$ which satisfies (i).

Furthermore if $\{C_1, \ldots, C_s\}$ is a set of orthogonal p-dimensional o.g.s. which generate $V_c(A)$ and $\{L_1, \ldots, L_s\}$ is the corresponding set of p-dimensional o.g.s. in R^{kp}, i.e. $E(C'_i \underline{x}) = L'_i \mu$ $i = 1, \ldots, s$, then

(4.2) $\underline{c}' = \underline{m}'_1 C'_1 + \underline{m}'_2 C'_2 + \ldots + \underline{m}'_s C'_s$

where $\underline{\ell}' = m'_1 L'_1 + \ldots + \underline{m}'_s L'_s$.

Proof: Using the notation of the theorem define p-dimensional o.g.s. $F_1, \ldots, F_{N-s}$ so that $\{C_1, \ldots, C_s, F_1, \ldots, F_{N-s}\}$ is an orthogonal set of o.g.s. which generates R^{Np}. Notice that $\{F_1, \ldots, F_{N-s}\}$ generates V(E). It is easy to show that $\{L_1, \ldots, L_s\}$ generates $V_r(A)$. Thus since $\underline{\ell} \in V_r(A)$ there exists a $\underline{d} \in R^{Np}$ so that $E\, \underline{d}'\, \underline{x} = \underline{\ell}'\, \underline{\mu}$. Furthermore there exists unique $p \times 1$ vectors $\underline{m}_1, \ldots, \underline{m}_s, \underline{n}_1, \ldots, \underline{n}_{N-s}$, so that

$$\underline{d}' = \underline{m}'_1 C'_1 + \ldots + \underline{m}'_s C'_s + \underline{n}'_1 F'_1 + \ldots + \underline{n}'_{N-s} F'_{N-s}$$

Let $\underline{c}' = \underline{m}'_1 C'_1 + \ldots + \underline{m}'_s C'_s$ and the result follows from the orthogonality of $V_c(A)$ and V(E) and lemma 4.2 (iii).

5. The "variance" of the best estimate of $L'\, \underline{\mu}$. In this section we shall develop results which will enable us to calculate the variance-covariance matrix of the best estimate of an estimable linear function $L'\, \underline{\mu}$. We shall first need to continue the investigation of the structure of the best estimate begun in corollary 4.1.1.

Recall that we have shown that there exists a unique o.g.s. C in $V_c(A)$ so that $C' \underline{x}$ is the best estimate of $L' \underline{\mu}$ and that (corollary 4.1.1) there is a p-dimensional o.g.s. in R^{kp}, Q, so that $C' = Q' A'$. Thus

(5.1) $\quad Q' A' A \underline{\mu} = E(Q' A' \underline{x}) = L' \underline{\mu}, \quad \forall \underline{\mu},$

which means that Q must satisfy

(5.2) $\quad A' A Q = L.$

Consistency of (5.2) follows from the fact that the estimability of $L' \underline{\mu}$ implies $L = A' B$ for B some p-dimensional o.g.s. in R^{Np} so that (5.2) is equivalent to

(5.3) $\quad A' A Q = A' B$

and the following Theorem (Bose's notes p. 52A) applied to the "column equations" in (5.3)

<u>Theorem 5.1</u> <u>The equations</u>

$$A' A \underline{q} = A' \underline{b}$$

<u>for $\underline{q}$ have a solution and $A \underline{q}$ is uniquely determined.</u>

However Theorem 5.1 gives us more than just consistency of (5.2). For suppose $\hat{\underline{\mu}}$ is a solution of

(5.4) $A' A \underline{\mu} = A' \underline{x}$

then best estimate of $L' \underline{\mu}$ is given by

(5.5) $Q' A' \underline{x} = Q' A' A \hat{\underline{\mu}} = L' \hat{\underline{\mu}}$.

Thus we have the following theorem.

Theorem 5.2 If $L' \underline{\mu}$ is estimable, the best estimate of $L' \underline{\mu}$ is $L' \hat{\underline{\mu}}$ where $\hat{\underline{\mu}}$ is any solution of (5.4). If $\underline{\ell}' \underline{\mu}$ is estimable, the best estimate of $\underline{\ell}' \underline{\mu}$ is $\underline{\ell}' \hat{\underline{\mu}}$.

Although (5.4) may not have a unique solution, $L' \hat{\underline{\mu}}$ will be unique no matter which solution, $\hat{\underline{\mu}}$, of (5.4) is used since $L' \hat{\underline{\mu}} = Q' A' A \hat{\underline{\mu}}$ and by theorem 5.1 $A \hat{\underline{\mu}}$ is uniquely determined. The equations (5.4) are called the normal equations. If we introduce the conditional inverse of $A' A$ into the discussion we get that a solution of (5.4) is given by

(5.6) $\hat{\underline{\mu}} = (A'A)^* A' \underline{x}$

for $(A'A)^*$ satisfying $(A'A) (A'A)^* (A'A) = A'A$, i.e. $(A'A)^*$ is a conditional inverse of $A'A$.

We are now in a position to state the following result.

Theorem 5.3 If $L' \underline{\mu}$ is estimable the variance covariance matrix of the best estimate of $L' \underline{\mu}$ is given by

(5.7) $VAR (L' \hat{\underline{\mu}}) = Q'A'AQ \Sigma_0 = L' Q \Sigma_0$ for Q satisfying (5.2).

Furthermore if $\{C_1, \ldots, C_s\}$ is a set of orthogonal p-dimensional o.g.s. which generate $V_c(A)$ and $\{L_1, \ldots, L_s\}$ is the corresponding set of p-dimensional o.g.s. in R^{kp} then we have the following corollary.

Corollary 5.3 *If $\underline{\ell} \in V_r(A)$ then the variance of the best estimate of $\underline{\ell}' \underline{\mu}$ is given by*

$$(5.8)\quad \mathrm{Var}(\underline{\ell}' \hat{\underline{\mu}}) = \sum_{i=1}^{s} \underline{m}'_i \, Q'_i A' A Q_i \, \Sigma_0 \, \underline{m}_i$$

$$= \sum_{i=1}^{s} \underline{m}'_i \, L'_i Q_i \, \Sigma_0 \, \underline{m}_i$$

$$= \sum_{i=1}^{s} \underline{\ell}'_i \, \underline{q}_i \, \underline{m}'_i \, \Sigma_0 \, \underline{m}_i$$

for $\underline{\ell}' = \underline{m}'_1 L'_1 + \ldots + \underline{m}'_s L'_s$ and $\underline{\ell}_i, \underline{q}_i$ the defining vectors of o.g.s. L_i, Q_i, $i = 1, \ldots, s$. *Here Q_i satisfies* $A' A Q_i = L_i$.

6. The Gauss-Markoff Theorem. For completeness we mention the Gauss-Markoff Theorem even though it contains nothing new in this case.

Theorem 6.1 (Gauss-Markoff). *Suppose $\hat{\underline{\mu}}$ is a vector which minimizes* $(\underline{x} - A\underline{\mu})'(\underline{x} - A\underline{\mu})$, *i.e. the squared length of the deviation of $\underline{x}$ from its mean. Then $L'\hat{\underline{\mu}}$ is the best estimate of $L'\underline{\mu}$ for $L'\underline{\mu}$ estimable and $\underline{\ell}'\hat{\underline{\mu}}$ is best estimate of $\underline{\ell}'\underline{\mu}$ for $\underline{\ell} \in V_r(A)$.*

7. Acknowledgment. The author here acknowledges the help and inspiration of the points of view expressed by Professor R. C. Bose and the late Professor S. N. Roy. Also discussions with Professors D. S. Burdick and W. M. O'Fallon here at Duke have been invaluable.

REFERENCES

[1] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. New York, Wiley.

[2] Bose, R. C. Unpublished notes on univariate analysis of variance. University of North Carolina, Chapel Hill.

[3] Roy, S. N. (1957). Some Aspects of Multivariate Analysis. New York, Wiley.

[4] Scheffe, H. (1959). The Analysis of Variance, New York, Wiley.

[5] Wedderburn (1934). Lectures on Matrices, New York, American Mathematical Society.